# Uncertainty of Geographic Information and Its Support in MADS

**Hong Shu[1,3], Stefano Spaccapietra[1], Christine Parent[2], Diana Quesada Sedas[1]**

[1]Database Laboratory (I&C-LBD) ,Swiss Federal Institute of Technology
1015 Lausanne, Switzerland
{Diana.QuesadaSedas, Stefano.Spaccapietra}@epfl.ch

[2]Institute of Inforge, University of Lausanne
1015 Lausanne, Switzerland
Christine.Parent@hec.unil.ch

[3]National Lab for Information Engineering in Surveying, Mapping and Remote Sensing
Wuhan University, 430079 Wuhan, China
Shu_Hong@hotmail.com

## Abstract
In this paper, we are aware of that geographic information uncertainty may arise from the complexity of the human-computing machine-earth system in general, and the differences among human cognition, computer representation and geographic reality in particular. To further clarify the nature of uncertainty, different types of uncertainty are hierarchically organized into a taxonomy of uncertainty. Particularly, we introduce how MADS is extended to support geographic spatio-temporal information uncertainty modeling. Primary uncertain spatio-temporal data types and uncertain spatio-temporal relationships are formally defined. The idea of multi-stage uncertainty resolution, including numerical indicators of data uncertainty at the stage of metadata model, is proposed.

## 1 Uncertainty Revisited
In Geoinformatics, it is widely acknowledged that uncertainty arises from limited computerization of infinitely complex geographic word. The limited computerization mainly refers to discrete representation, finite levels of detail, incomplete data collection, deficient knowledge, etc.. It is often said that uncertainty is an inherent property of GIS data or geographic phenomena. However, by the inherent property what does it mean? Undoubtedly, we need a context of assigning uncertainty with an exact meaning. To this end, we propose the human-computing machine-earth system as this type of context as in Figure 1.

At the top level, we consider human being, computing machine and the earth as three highly-abstracted entities which are mutually related and interacted for constructing a real word, or say, the human-computing machine-earth system. As far as the interaction is concerned, we have human-environment interaction, human-machine interaction, and computer-virtualized or augmented reality. Roughly, geographic information science could be known as an interdisciplinary science of cognitive science, geography and computer science.

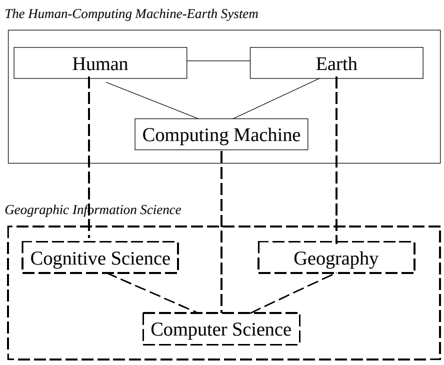


**Figure 1. The Human-Computing Machine-Earth system and geographic information science**

From this perspective, it seems easy for us to work out the nature of uncertainty. As shown in Figure 2, uncertainty arises from the complexity of the human-computing machine-earth system. More precisely, uncertainty arises from the differences among human being, computing machine and the earth as well as the differences within human being or computing machine or the earth. Geographic information uncertainty reflects the richness of geographic states, the inability of human cognition, and the limited computing capacity of machine.

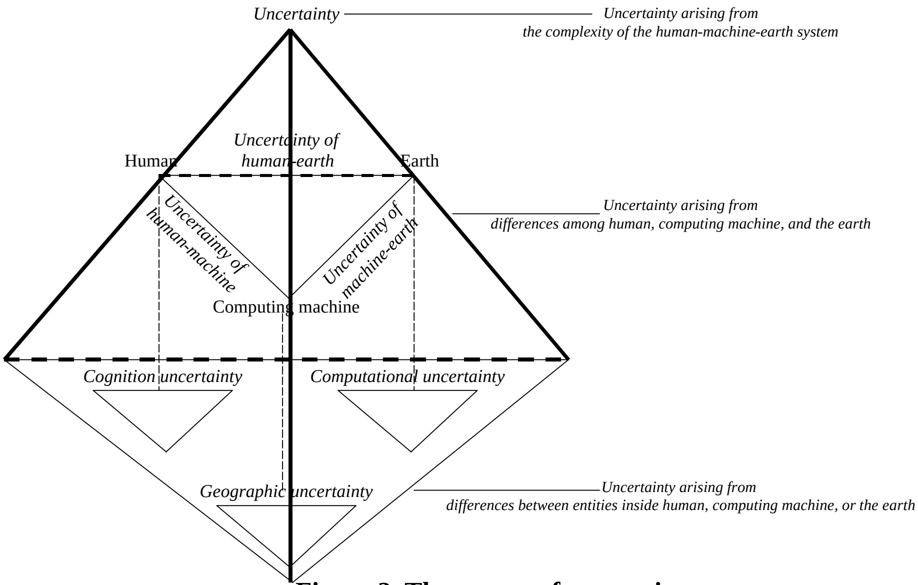


**Figure 2. The nature of uncertainty**

From another angle, it is the differences among geographic reality, computer representation, and human cognition that make it possible for human being, computing machine and the earth to interact and correlate with each other. Increase of interaction and correlation implies decrease of differences between entities. When geographic reality, computer representation, and human cognition are highly related or consistent, uncertainty is somewhat removed. In this sense, the truth is achieved through the coherency of entities.

## 2 Taxonomy of Uncertainty

To extend our basic ideas about the nature of uncertainty, we propose a taxonomy of uncertainty in Figure 3. The taxonomy of uncertainty is intended to put different types of uncertainty into a unified framework, which will serve as a specification of conceptualization of uncertainty. In practice, the taxonomy of uncertainty is useful for developing an integrated uncertainty model or examining transformations of different types of uncertainty.

The generic uncertainty is divided into uncertainty of entities and uncertainty of human-machine-earth relations. Uncertainty of entities is implied by the differences among entities inside human cognition or computing machine or the earth. Uncertainty of human-machine-earth relations arises from the differences among cognitive, computational and geographic entities. Moreover, uncertainty of cognitive entities is mainly divided into perceptual uncertainty, memory uncertainty and thinking uncertainty. Typically, among perceptual uncertainty is visual uncertainty, such as visual haze. The lack of knowledge (ignorance) is renown as memory uncertainty. Insolvability of problem (undecidability) falls into thinking uncertainty. Non-classical logical reasoning, including modal logic, multi-valued logic, non-monotonic logic, qualitative calculus, probabilistic and possibilistic logic, may be put into thinking uncertainty.

As for uncertainty of human-machine-earth relations, it is further divided into inaccuracy, incompleteness, inconsistency and imprecision. Inaccuracy, also called error, refers to some deviation of measurement value from the true value. The slight deviation of the measurement value from the true value is named approximation, and the serious deviation of the measurement value from the true value is named incorrectness or the wrong. Incompleteness means missing of some values, frequently interpreted as partial computational and cognitive description of geographic phenomena. Inconsistency means that, for the same geographic entity, there exists several different computational and cognitive statements. In degree of inconsistency, there exists representational diversity (i.e., conflict), semantic mismatch (i.e., incoherence) and semantic contradiction (i.e., invalidity). Imprecision refers to the degree of exactness of computational and cognitive values, which is closely related to the resolution. Imprecision with a low resolution of values is named non-specificity. In non-specificity, the true value falls into an interval of possible values, e.g., disjunction or negation of possible values. Imprecision with a lower resolution of values is named ambiguity or confusion. In ambiguity, it is difficult for us to find out such an interval into which the true value falls. Imprecision with the much lower resolution of values is named vagueness or fuzziness. By fuzziness it means the true value is gradually changing from the false to the truth. That is, no sharply defined boundary is given between the false and the truth.

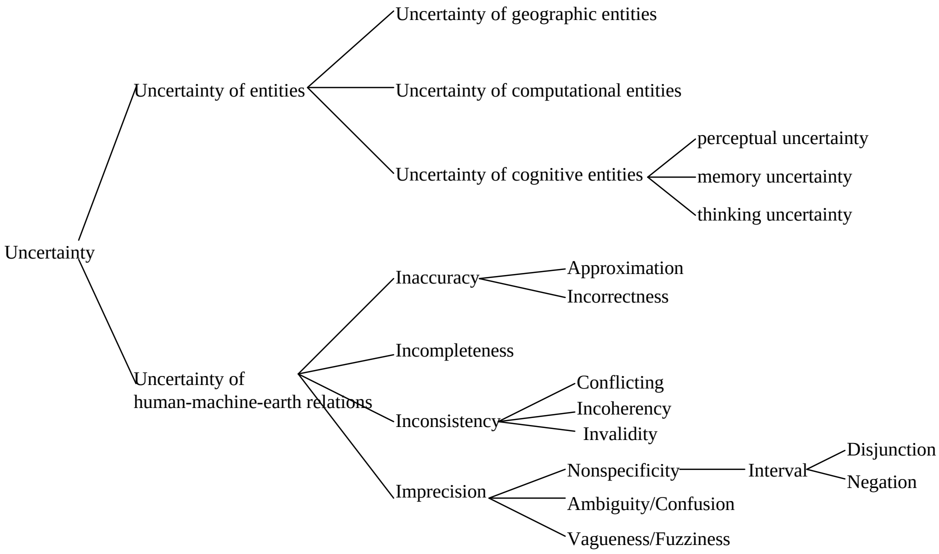


**Figure 3. A taxonomy of uncertainty**

# 3 Dealing with Geographic Information Uncertainty in MADS

MADS is a GIS conceptual data model with characteristics of 1) spatio-temporal conceptual modeling, 2) ODMG-conformant, 3) application- and resolution-adapted representation, 4) visually aid schema design (MADS, 1997; Parent C., Spaccapietra S., and Zimanyi E., 1999; Parent C., Spaccapietra S., and Zimanyi E., 2000). In this section, we present how geographic information uncertainty modeling is supported in MADS. In particular, primary uncertain spatio-temporal data types and uncertain spatio-temporal relationships are formally defined. The idea of multi-stage uncertainty resolution, including numerical indicators of data uncertainty implemented at the metadata stage, is proposed.

## 3.1 Primary Uncertain Spatio-temporal Data Types

As shown in Figure 4, uncertainty of a geographic entity can be modeled through uncertainty of its geospatial, temporal and thematic attributes.

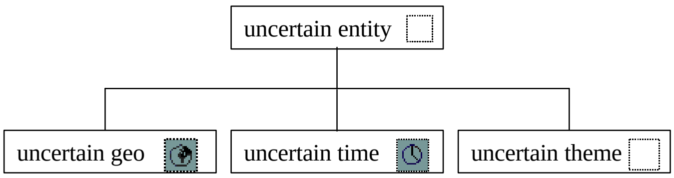


**Figure 4. The uncertainty of an entity and its attributes of the space, time and theme**

A hierarchical system of uncertainty spatial data types is designed as in Figure 5. That is, {uncertain geo}, {random geo and fuzzy geo}, {primary uncertainty geospatial data types}. The formal definitions of these primary data types are given as follows.

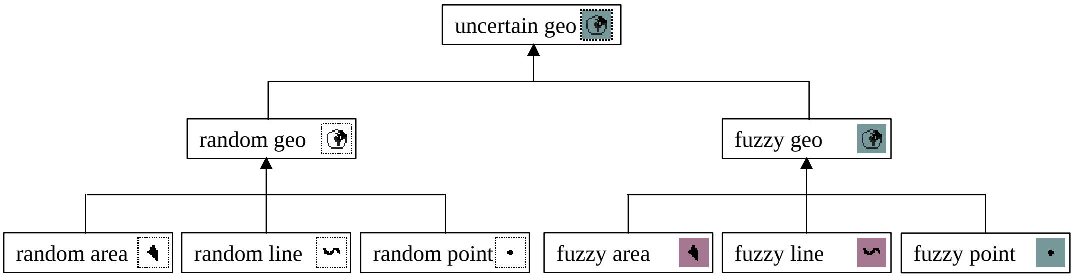


**Figure 5. Uncertainty geospatial data types**

It is assumed that Geographic space, denoted $GSpace$, is a two-dimensional real Euclidean space ( $IR^2$ ). Geometrically, $GSpace$ is a set of geographic coordinate points. Computationally, the geographic entity is modeled with a geographic object. The geographic object, denoted $GObject$, is a subset of $GSpace$.

**Definition 1** ( $random\ \ point$ ). A random point is used for spatially modeling a point-like geographic entity stochastically occurred. A random point is a coordinate point associated with its probability, denoted $p_r((x,y),P)$ . The domain of data type " $random\ \ point$ " is a set of random points. $random\ \ point$ functionally performs a mapping from $GSpace$ to a unit interval of probability, i.e., $random\ \ point : GSpace \rightarrow [0,1]$ .

For example, with a GPS receiver, the coordinate point $(10.23cm, 32.20cm)$ will be collected with a probability of *0.9* at next moment. It is modeled as a random point $p_r((10.23cm, 32.20cm), 0.9)$ .

**Definition 2** ( $random\ \ line$ ). A random line is used for spatially modeling a linear geographic entity stochastically occurred. Random line can be modeled in holistic or reductionistic way.

In a holistic way, the random line is modeled as a deterministic line associated with its probability of occurrence, denoted $l_r(l,P)$ . The domain of data type " $random\ \ line$ " is a set of random lines. Let a set of deterministic lines be a space of lines, denoted $LSpace$ . $random\ \ line$ functionally performs a mapping from $LSpace$ to a unit interval of probability, i.e., $random\ \ line : LSpace \rightarrow [0,1]$ .

For example, from the Lausanne train station CFF to the university EPFL, we can take bus or metro. The route line of CFF to EPFL can be modeled with a random line, bus line or metro line, i.e., $routeCFFtoEPFL\{(busline, 0.5), (metroline, 0.5)\}$ .

Alternatively, in a reductionistic way, the random line is modeled as a set of random points of which the random line is composed, denoted $l_r(p_{r1}, p_{r2}, ..., p_{rn})$ . For brevity, it is assumed that random points on the line are independent and in accordance with the same probability distribution.

For example, a GPS-instrumented post car goes from the main post office to EPFL every day. On the electronic map, the post car route is modeled with a sequence of random points sampled by GPS.

**Definition 3** ( $random\ \ area$ ). A random area is used for spatially modeling an area geographic entity stochastically occurred. Analogous to $random\ \ line$ , random area can be modeled in holistic or reductionistic way.

In a holistic way, the random area is modeled as a deterministic area associated with its probability of occurrence, denoted $a_r(a,P)$ . The domain of data type " $random\ \ area$ " is a set of random areas. Let a set of deterministic areas be a space of areas, denoted $ASpace$. $random\ \ area$ functionally performs a mapping from $ASpace$ to a unit interval of probability, i.e., $random\ \ area : ASpace \rightarrow [0,1]$.

For example, consider your car as an area-like object on the large-scale map, the car may be parked on the roadside or on the spot. Assume that the car is parked on the spot with a probability of 0.95. Then the car is modeled with a random area.

Alternatively, in a reductionistic way, an area is approximately represented with its boundary. Thus, the random area can be modeled with its indeterministic boundary, i.e., a random polyline. In this case, the random area is denoted $a_r(l_r)$.

For example, Geneva Lake is stochastically changing as lake water rising up and falling down. The random area, e.g., Geneva Lake, can be identified with its random boundary.

For fuzzy geographic objects, we have primary fuzzy data types of fuzzy point, fuzzy line and fuzzy area. In our project, the possibility theory is chosen for fuzzy object modeling. The possibility theory is derived from fuzzy set theory. The possibility of the object being taken under the fuzzy constraint is numerically equal to its grade of membership in a fuzzy set. Mathematically, a fuzzy constraint, denoted $FR$, is represented with a fuzzy set.

**Definition 4** ( $fuzzy\ \ point$ ). A fuzzy point is used for spatially modeling a point-like vaguely defined geographic entity. A fuzzy point is a coordinate point associated with its possibility, denoted $p_f((x,y),\Pi)$. The domain of data type " $fuzzy\ \ point$ " is a set of fuzzy points. Under the fuzzy constraint, the possibility distribution of $fuzzy\ \ point$ performs a mapping from $GSpace$ to a unit interval of possibility, that is, $p^{p}_{FR} : GSpace \rightarrow [0,1]$.

Note that the unit intervals of probability and possibility are semantically different. The unit interval [0,1] in the definition of a fuzzy object is the range of the possibility of object being taken. However, the unit interval [0,1] in the definition of a random object is the range of the probability of object' s occurrence.

For example, the Lausanne train station CFF at $(x_0, y_0)$ can be modeled as a point on the small-scale map. The possibility of the city center of Lausanne being at train station $(x_0, y_0)$ is 0.85. Thus, we have the fuzzy point, $CFF_{"city\ \ center"}((x,y),0.85)$.

**Definition 5** ( $fuzzy\ \ line$ ). A fuzzy line is used for spatially modeling a linear vaguely defined geographic entity. Like $random\ \ line$, fuzzy line can be modeled in holistic or reductionistic way.

In a holistic way, the fuzzy line is modeled as a deterministic line associated with the possibility of its being taken under a fuzzy constraint, denoted $l_f(l,\Pi)$. The domain of data type " $fuzzy\ \ line$ " is a set of fuzzy lines. Assume that $LSpace$ is a set of deterministic lines. Under a fuzzy constraint $FR$, the possibility distribution of $fuzzy\ \ line$ functionally performs a mapping from $LSpace$ to a unit interval of the fuzzy, i.e., $p^{l}_{FR} : LSpace \rightarrow [0,1]$.

For example, we can take bus or metro from the Lausanne train station CFF to the university EPFL. For travelers who prefer to walk a scenic road, the possibility of metro line being a scenic road is 0.72. In this case, the metro line from CFF to EPFL can be modeled as a fuzzy line, $metroCFFtoEPFL_{"scenic\ \ road"}(metroCFFtoEPFL,0.72)$.

Alternatively, in a reductionistic way, the fuzzy line is modeled as a set of fuzzy points of which the fuzzy line is composed, denoted $l_f(p_{f1}, p_{f2},..., p_{fn})$. For brevity, it is assumed that fuzzy points on the line are independent and in accordance with the same possibility distribution.

For example, assume that the metro line from Lausanne train station CFF to the university EPFL is interpreted from an aerial photo taken over the sky of Lausanne city. On a small-scale imagery map, Lausanne train station CFF is interpreted as a fuzzy point where the metro line starts. In this case, the metro line from CFF to EPFL is modeled as a fuzzy line starting from a fuzzy point "Lausanne train station CFF".

**Definition 6** ( $fuzzy \quad area$ ). A fuzzy area is used for spatially modeling a vaguely defined geographic entity. Likewise, fuzzy area can be modeled in holistic or reductionistic way.

In a holistic way, the fuzzy area is modeled as a series of ë-level cut sets of the fuzzy set, denoted $a_f(\{a_l, l \in [0,1]\})$. A ë-level cut set of the fuzzy set is the set of all elements whose grades of membership are greater than or equal to **"**ë**"**. When ë takes "1", the cut set is interpreted as the interior of the fuzzy set. When ë takes "0", the cut set is interpreted as the exterior of the fuzzy set. When ë takes a certain value from the interval **(0,1**), all boundaries of cut sets constitute a fuzzy boundary of the fuzzy set. The domain of data type " $fuzzy \quad area$ " is a set of fuzzy areas. Let a set of deterministic areas be a space of areas or a set of ë-level cut sets, denoted $ASpace$. $fuzzy \quad area$ functionally performs a mapping from $ASpace$ to a unit interval of possibility, i.e., $fuzzy \quad area : ASpace \to [0,1]$.

For example, consider Lausanne city as a fuzzy area on the large-scale map, city area is gradually growing from the city center into the suburb in degree of urbanization. Here the degree of urbanization can be modeled as a level of fuzzy area "city". Thus, we have a fuzzy city, $Lausanne\ city(\{the\ city\ area, the\ degree\ of\ urbanization\})$.

Alternatively, in a reductionistic way, an area is approximately represented with its boundary. Thus, the fuzzy area can be modeled with its vaguely-defined boundary, i.e., a fuzzy polyline. In this case, the fuzzy area is denoted $a_f(l_f)$.

For example, on the city map of Lausanne city, the commercial area is vaguely defined. It can be modeled approximately with a fuzzy polyline.

Similarly, uncertain time data types are divided into random time and fuzzy time as in Figure 6. Moreover, random time is divided into random instant and random interval, and fuzzy time is divided into fuzzy instant and fuzzy interval. Definitions of uncertain time data types are given as follows.

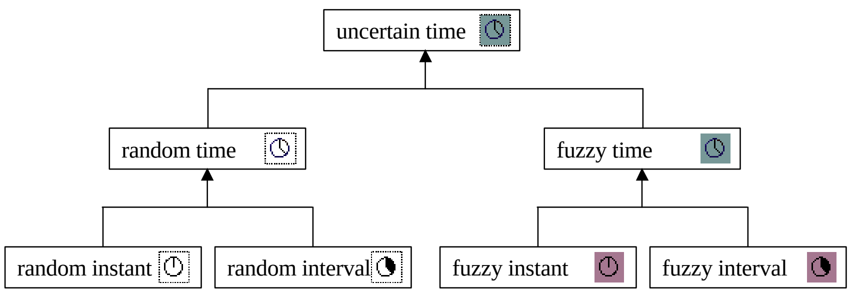


**Figure 6. Uncertainty temporal data types**

Time is modeled as a linear Euclidean space ( $IR$ ), denoted $TSpace$. $TSpace$ is a set of time coordinate points. A time coordinate is granulated with a chronon. Chronon is a minimum unit of time implemented in the computer (Dyreson Curtis E. and Richard T. Snodgrass, 1993). Determined by the computer functionality, chronon may be millisecond, second, minute, hour, day, month, year, etc. The size of a valid time instant may be greater than or smaller than a chronon. For brevity, it is assumed that a valid time instant is of the same size as a chronon.

Source of time uncertainty may be: 1) Imprecise dating techniques, e.g., the millisecond is indeterministic by watch. 2) Planning time, e.g., completed time of a new building is uncertain by virtue of unpredicted factors. 3) Forgotten time, e.g., we can' t verify the construction time of an ancient building due to the lack of historical documents.

**Definition 7** ( *random instant* ). A random instant is used for temporally modeling a geographic event stochastically occurred. A random instant is a time coordinate point associated with its probability, denoted $t_r(t, P)$ . The domain of data type " *random instant* " is a set of random instants. *random instant* functionally performs a mapping from *TSpace* to a unit interval of probability, i.e., $random\ \ instant : TSpace \to [0,1]$ .

For example, the travel bus arrives at the foot of the mountain at 10:23 with a probability of *0.85*. This is modeled as a random instant $t_r(10:23, 0.85)$.

**Definition 8** ( *random interval* ). A random interval is used for temporally modeling a geographic process stochastically occurred. A random interval may be modeled holistically or reductionistically.

In a holistic way, the random interval is modeled as a deterministic interval associated with its probability of occurrence, denoted $i_r(i, P)$ . The domain of data type " *random interval* " is a set of random intervals. Let a set of deterministic intervals be a space of intervals, denoted *ISpace* . *random interval* functionally performs a mapping from *ISpace* to a unit interval of probability, i.e., $random\ \ interval : ISpace \to [0,1]$ .

For example, tomorrow it is raining from 9:15 until 13:31 with a probability of 0.56. This is modeled as $raining([9:15, 13:31], 0.56)$.

Alternatively, in a reductionistic way, the random interval is modeled as a set of random instants of which the random interval is composed, denoted $i_r(t_{r1}, t_{r2}, ..., t_{rn})$ . For brevity, it is assumed that random instants in the interval are independent and in accordance with the same probability distribution.

For example, the rainy week will begin from Dec. 2 until Dec. 7 with probabilities of 0.6, 0.7, 0.8, 0.9, 0.7,0.5 respectively. This is modeled as $rainyweek((Dec.2,0.6),(Dec.3,0.7),(Dec.4,0.8),(Dec.5,0.9),(Dec.6,0.7),(Dec.0.5))$.

**Definition 9** ( *fuzzy instant* ). A fuzzy instant is used for temporally modeling a vaguely defined geographic event. A fuzzy instant is a time coordinate point associated with its possibility, denoted $t_f(t, \Pi)$ . The domain of data type " *fuzzy instant* " is a set of fuzzy instants. Under a fuzzy constraint, the possibility distribution of *fuzzy instant* functionally performs a mapping from *TSpace* to a unit interval of possibility, i.e., $p^{t}_{FR} : TSpace \to [0,1]$ .

For example, the travel bus arrives at the foot of the mountain at around 10:23, more precisely at 10:23 with a possibility of 0.93. It is explained that the possibility of bus arrival time being 10:23 is 0.93.

**Definition 10** ( *fuzzy interval* ). A random interval is used for temporally modeling a vaguely defined geographic process. Likewise, a fuzzy interval may be modeled holistically or reductionistically.

In a holistic way, the fuzzy interval is modeled as a deterministic interval associated with its possibility of occurrence, denoted $i_f(i,\Pi)$. The domain of data type " *fuzzy interval* " is a set of fuzzy intervals. Assume that $ISpace$ is a set of deterministic intervals. Under a fuzzy constraint $FR$, the possibility distribution of *fuzzy interval* functionally performs a mapping from $ISpace$ to a unit interval of the fuzzy, i.e., $p^{i}_{FR} : ISpace \rightarrow [0,1]$.

For example, a marathon running day in Lausanne is scheduled on Nov. 1 with a possibility of 0.7. It means that, from 9:00 until 17:00 of Nov. 1, a marathon running race will be held in Lausanne with a possibility of 0.7. It is denoted *marathon([9 : 00, Nov.1,17 : 00, Nov.1],0.7)*.

Alternatively, in a reductionistic way, the fuzzy interval is modeled as a set of fuzzy instants of which the fuzzy interval is composed, denoted $i_f(t_{f1}, t_{f2},...,, t_{fn})$. For brevity, it is assumed that fuzzy instants in the interval are independent and in accordance with the same possibility distribution.

For example, a cloudy week will appear from Nov. 11 until Nov. 16 with the possibility of 0.2, 0.3, 0.6, 0.8, 0.9,0.1 respectively. That means that the possibilities of a cloudy week being the days of Nov. 11 until Nov. 16 are 0.2, 0.3, 0.6, 0.8, 0.9, 0.1 respectively. This is modeled as:
*cloudyweek((Nov.11,0.2),(Nov.12,0.3),(Nov.13,0.6),(Nov.14,0.8),(Nov.15,0.9),(Nov.16,0.1))*

In this section, a geographic entity is actually modeled as a random or fuzzy spatio-temporal variable. In the computer, the parameters of probability or possibility functions are implemented as metadata items.

### 3.2 Numerical Indicators of Data Uncertainty

Basically, probabilistic and possibilistic distributed entities are mainly modeled by means of abstract data types, e.g., uncertain spatio-temporal data types. Alternatively, we provide users with some statistics of data uncertainty, called numerical indicators of data uncertainty. In the computer, numerical indicators of data uncertainty are implemented with metadata items.

For random spatio-temporal variables, standard deviation or standard error is chosen as numerical indicators of data uncertainty. Typically, for spatio-temporal variables or objects, there are the error circle for random points, the error band for random lines and random areas. For more details, readers are referred to (Zhang Jingxiong and Michael F. Goodchild, 2002). For fuzzy spatio-temporal variables or objects, fuzzy measures can be taken as numerical indicators of data uncertainty. For thematic variables, some statistics of contingency table are often taken as numerical indicators of data uncertainty. One of thematic data uncertainty indicators is PCC (percentage of correctly classified categories derived from an image classification error matrix).

A comprehensive framework for uncertainty model is in the form of:

*Entity*
*(*
*Identifier,*
*Uncertainty model of spatial, temporal and thematic attributes,*
*Numerical indicators of uncertainty of spatial, temporal and thematic data,*
*Operation of data uncertainty,*
*)*

In this framework, the identifier of the entity is an unambiguous name or a unique alphanumerical string. To a large extent, uncertainty model of spatial, temporal and thematic attributes are probability or possibility distribution of attribute values of the entity, which are implemented by means of uncertainty data types. Numerical indicators of uncertainty of spatial, temporal and thematic data for the entity are a set of pairs (indicator name, indicator value), which are implemented as metadata. As an example, uncertainty modeling of point, line, and area objects are illustrated in Tables 1-3. For brevity, only random data types and related uncertainty indicators are presented here.

**Table 1. An example for uncertainty modeling of a point**

| *Data model* | | | Metadata model | |
|---|---|---|---|---|
| *Identifier* | *Uncertainty model of spatial attributes (Coordinate point, probability)* | | Numerical indicators of uncertainty of spatial data (Indicator name, indicator value) | |
| *Metro stop "Ouchy"* | *(101.0m, 234.2m)* | *0.78* | Circle standard deviation of the metro stop | 0.07m |
| | *…* | *…* | | |
| | *(101.5m, 233.8m)* | *0.81* | | |

**Table 2. An example for uncertainty modeling of a line**

| *Data model (line as a set of points)* | | | Metadata model | |
|---|---|---|---|---|
| *Identifier* | *Uncertainty model of spatial attributes (Coordinate point, probability)* | | Numerical indicators of uncertainty of spatial data (Indicator name, indicator value) | |
| *Edouard Dapples Street* | *(33.31m, 25.02m)* | *0.89* | å-band width of the street | *0.06m* |
| | *(33.02m, 26.23m)* | *0.91* | | |
| | *…* | *…* | | |
| | *(89.33m, 68.81m)* | *0.90* | | |

**Table 3. An example for uncertainty modeling of an area**

| *Data model(area as a set of lines, and line as a set of points)* | | | Metadata model | |
|---|---|---|---|---|
| *Identifier* | *Uncertainty model of spatial attributes (Coordinate point, probability)* | | Numerical indicators of uncertainty of spatial data (Indicator name, indicator value) | |
| *Lake Leman* | *(6325.21m, 3289.52m)* | *0.89* | å-band width of the waterfront line | 0.11m |
| | *(6325.79m, 3290.01m)* | *0.88* | | |
| | *…* | *…* | | |
| | *(6325.21m, 3289.52m)* | *0.89* | | |

### 3.3 Uncertain Spatio-temporal Relationships

In theory, Clementini (Clementini Eliseo, Paolino Di Felice, and Peter van Oosterom, 1993) proposed 5 spatial topological relationships as a minimum set of topological relationships. (Allen J. F., 1983) proposed 13 temporal relationships mixing temporal topological relationships with temporal ordering relationships, which have gained wide popularity of application in natural language processing and others. Our work is based on Clementini's spatial topological relationships and Allen's temporal relationships.

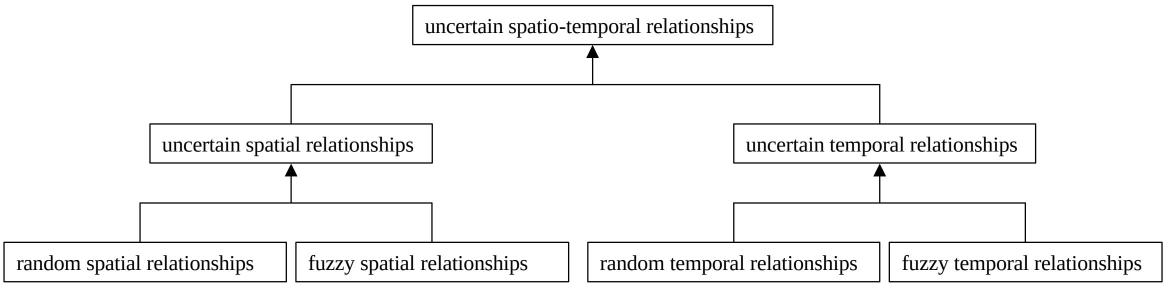


**Figure 7. Uncertainty spatio-temporal relationships**

Likewise, uncertain spatio-temporal relationships can be modeled in holistic or reductionlistic way. In a reductionlistic way, some functional operators or predicates are provided to users for automatically extracting possible spatio-temporal relationships from uncertain spatio-temporal objects. In other words, uncertainty of objects is prorogated into uncertainty of relationships between objects. Since spatio-temporal objects are of various dimensions and in a complex structure, extraction of uncertain spatio-temporal relationships need to employ tedious algorithms of computational geometry and topology. Examination of possible spatial topological relationships existing between indeterminate area (region) objects (or say, objects with indeterminate/broad boundaries) has been made in (Cohn A. G. and N. M. Gotts, 1996), (Clementini Eliseo and Paolino Di Felice, 1996). As a complementary result, here we only discuss the modeling of uncertain spatio-temporal relationships in a holistic way. The formal definitions of uncertain spatio-temporal relationships are given as follows.

Let space of spatial topological relationships be a set of deterministic spatial topological relationships, denoted $SRSpace$. In the viewpoint of Clementini, we have $SRSpace = \{ s\_disjoint, s\_touch, s\_overlap, s\_contain, s\_equal \}$.

Let space of temporal relationships be a set of deterministic temporal relationships, denoted $TRSpace$. In the viewpoint of Allen, we have *TRSpace = {t_before, t_meet, t_overlap, t_start, t_during, t_finish, t_equal, t_finished, t_contain, t_started, t_overlapped, t_met, t_after}*. It is easily found that Allen's temporal relationships is a temporal ordering refinement of temporal topological relationships *{t_disjoint, t_touch, t_overlap, t_contain, t_equal}*). For example, topological relationship "*t_disjoint*" is distinguished into "*t_before*" and "*t_after*" disjoints.

**Definition 11** (*random spatial relationships*). A random spatial relationship is used for modeling spatial relationships between geographic entities stochastically occurred. A random spatial relationship is a deterministic spatial relationship associated with its probability, denoted $sr_r(sr, P)$. The probability function of *random spatial relationships* functionally performs a mapping from $SRSpace$ to a unit interval of probability, i.e., $random\ \ spatial\ \ relationships : SRSpace \rightarrow [0,1]$.

**Definition 12** (*fuzzy spatial relationships*). A fuzzy spatial relationship is used for modeling spatial relationships between vaguely defined spatial objects. A fuzzy spatial relationship is a deterministic spatial relationship associated with its possibility, denoted $sr_f(sr, \Pi)$. Under a fuzzy constraint $FR$, the possibility distribution of *fuzzy spatial relationships* functionally performs a mapping from $SRSpace$ to a unit interval of possibility, i.e., $p_{FR}^{sr} : SRSpace \rightarrow [0,1]$.

**Definition 13** (*random temporal relationships*). A random temporal relationship is used for modeling temporal relationships between geographic events stochastically occurred. A random temporal relationship is a deterministic temporal relationship associated with its probability, denoted $tr_r(tr, P)$. The probability function of *random temporal relationships* functionally performs a mapping from $TRSpace$ to a unit interval of probability, i.e., $random\ \ temporal\ \ relationships : TRSpace \rightarrow [0,1]$.

**Definition 14** (*fuzzy temporal relationships*). A fuzzy temporal relationship is used for modeling temporal relationships between vaguely defined temporal objects. A fuzzy temporal relationship is a deterministic temporal relationship associated with its possibility, denoted $tr_f(tr, \Pi)$. Under a fuzzy constraint $FR$, the possibility distribution of *fuzzy temporal relationships* functionally performs a mapping from $TRSpace$ to a unit interval of possibility, i.e., $p_{FR}^{tr} : TRSpace \rightarrow [0,1]$.

It is evident that how to determine a probability distribution or possibility distribution function is crucial for randomizing and fuzzifying spatio-temporal relationships. Two simplified methods of computation are taken here. The first one is that, joint probability of two objects involved in relationship computation is thought of as probability of the relationship (see Eq. 1), and membership associated with intersection of two fuzzy objects (subset of points) is thought of as membership (or possibility) of the relationship (see Eq. 2). In both cases, we make an assumption that objects involved in relationship computation are independent.

$$P(relationship) = P(AB) = P(A) * P(B) \qquad (Eq.1)$$

$$\Pi(relationship) = m_{A \cap B}(relationship) = \min\{m_A(relationship), m_B(relationship)\} \qquad (Eq.\ 2)$$

In second method, the probability and possibility of the relationship are computed through analysis of structures of $SRSpace$ and $TRSpace$. Specify some type of changes, such as moving, enlarging, etc., conceptual distance between two relationships is the number of nodes of relationships-connected path in a conceptual neighborhood graph of $SRSpace$ or $TRSpace$. It is observed that conceptual distance functions implicitly reflect structures of $SRSpace$ and $TRSpace$. Thus, the probability and possibility functions of relationships may be derived from some conceptual distance functions. Related studies have been somewhat conducted in (Guesgen Hans W., 2001).

### 3.4 Multi-stage Uncertainty Resolution

With a reference to our taxonomy of uncertainty, geographic information uncertainty modeling in MADS is supported at multiple stages of data model, metadata model, and interactive software architecture. At the stage of data model, uncertain data types and relationships are used. Besides, two special values of uncertainty, “NULL” and “Now”, are resolved in a particular way.

“NULL” have two semantics, “unknown” and “inapplicable” (Clifford James, Curtis E. Dyreson, Richard T. Snodgrass, Tomás Isakowitz, and Christian S. Jensen, 1995). “unknown” means that, for an attribute, indeed there exists a value of this attribute but we are ignorant about its specific value at present. “inapplicable” means that it doesn’t exist any valid value for this attribute at all. In MADS, “unknown” is modeled with random data types, and “inapplicable” modeled with fuzzy data types. For example, it is unknown that tomorrow’s road traffic is jammed or free. It makes sense of assigning an equal probability to alternative status of tomorrow’s road traffic, i.e., roadtraffic{(jammed, 0.5), (free, 0.5)}. In a layer-based GIS, sometimes a point object and a line object are put in the same layer and share the same attribute description table. For example, for a point object, the attribute of “length” is apparently inapplicable, of taking the value of “NULL”. In MADS, we can model the value of length of a point object with ([min, max], 0). [min, max] is the range of lengths of all line objects involved in that layer. “0” is the possibility of the point taking any length value in [min, max]. “Now” is a special value of uncertainty in temporal databases. In the real word, “Now” refers to current time. In databases, “Now” means “until changed”. In MADS, “Now” is implemented as a variable, whose value is dynamically computed by a predefined function.

To make databases plausible, some mechanisms of maintaining data validity and completeness and consistency are needed. For human decision-making, it is important to reduce uncertainty of invalidity, incompleteness and inconsistency. Thus, some plausibility check rules are defined in databases. Traditional database management techniques are still effective for data plausibility control.

At the stage of metadata model, basic numerical indicators of uncertainty are specified. Values of indicators of data uncertainty can be computed statically before database creation, or be computed dynamically through some simulation methods, such as Monte Carlo simulation.

At the stage of interactive software architecture, computers cooperatively work with users for problem solving. Users’ task/requirement analysis and users’ control on the computer interface are carefully examined as well as computational models and computer feedbacks on users’ input. As explained before, uncertainty is rooted in the differences among realistic, computational and human cognitive models. In this sense, uncertainty modeling has inherently transformed into effective human-computer interaction. In this aspect, a large amount of work about interactive software architecture design can be used indirectly for the purpose of uncertainty modeling.

## 4 Related Work

Uncertainty modeling is a long-standing research issue in the areas of probability statistics and fuzzy mathematics, artificial intelligence and databases, geostatistics and error theory, etc. Over a few decades, different theories of uncertainty have been posed, but it is still lack of a comprehensive theory of uncertainty. Until very recently, the study of generalized uncertainty has increasingly attracted attention of scientists. Among them is George J. Klir (Klir G. J. and Wierman, 1997) who has come up with generic principles of uncertainty and generic uncertainty measures based on information and decision theory. Also, a few taxonomies of uncertainty have been theoretically proposed. Through our analysis, the taxonomies of uncertainty proposed in (Smets Philippe, 1991, 1996), (Bonissone Piero P. and Richard M. Tong, 1985), (Bosc Patrick and Henri Prade, 1996) result from semantic examination of uncertainty-formalizing mathematics, while Smithson largely concentrates on uncertainty concerning social decision (Smithson M.J., 1989). Smithson’s taxonomy of uncertainty is classified into two categories, i.e., status of ignorance (error) and act of ignorance (irrelevance). Indeed, it is difficult to make a comprehensive evaluation about soundness and completeness of these uncertainty taxonomies. In contrast, our taxonomy of uncertainty given in the context of the human-computing machine-earth system is intentionally served as a generic framework for geographic information uncertainty modeling.

In uncertainty spatio-temporal data management, Markus Schneider put forward a set of fuzzy spatial data types (Schneider Markus, 1999), where spatial objects are modeled as a set of points and fuzziness modeled by means of memberships. This is somewhat different from our methods of holistic spatial objects modeling and possibility-based fuzziness modeling. Holistic spatial objects modeling is complementary to reductionistic modeling of spatial objects being the sets of points. And possibility-based fuzziness modeling is advantageous in computability over membership-based fuzziness modeling.

For uncertain spatio-temporal relationships modeling, Guesgen suggested to fuzzify spatio-temporal relationship with characteristic functions (membership functions). In Guesgen’s work, traditional spatial and temporal relationships are thought of as a set of deterministic spatio-temporal relationships. Essentially, this is our holistic idea of uncertain spatio-temporal relationships modeling. By using 9-intersection algebraic theory and RCC logical method, Cohn A. G. and N. M. Gotts, Clementini Eliseo and Paolino Di Felice have addressed the possible set of spatial relationships existing between vague spatial regions. Basically, this is our reductionlistic idea of uncertain spatial topological relationships modeling, since uncertain spatial relationships are reduced to two related vague spatial objects. Thus, our solution could be said to be a combination of holistic and reductionlistic ideas of uncertainty spatio-temporal relationships modeling.

## 5 Conclusion

Two parts of work have been carried out in this paper. Firstly, we attempt to explore the nature of uncertainty in a broad sense. It is stated that geographic information uncertainty may arise from the complexity of the human-computing machine-earth system in general, and the differences among human cognition, computer representation and geographic reality in particular. In this view, a taxonomy of uncertainty is proposed for further clarifying our understanding about uncertainty.

Secondly, we have studied how to deal with uncertainty in MADS. Inspired by the theories of stochastic and fuzzy geometry (topology), methods of probabilistic, fuzzy and statistic databases, primary uncertain spatio-temporal data types and spatio-temporal relationships are formally defined. Innovatively, an idea of multi-stage uncertainty resolution is proposed.

However, our proposed uncertainty model is still at an initial stage. For example, uncertainty behavior modeling, including uncertainty propagation, is less discussed. All uncertainty data types are defined based on the assumption of simple spatial objects at a single granularity. The assumption of probabilistic and possibilistic independent objects with the same probability and possibility distribution has been made throughout this paper. This will certainly restrain MADS from flexibly modeling geographic spatio-temporal information uncertainty. Last but not least, it is noteworthy that our proposed taxonomy of uncertainty may be incomplete in theory. This is related to our ideas on the one hand, and is influenced by the statements of natural language on the other. It is well known that natural language is of sense ambiguity. Moreover, philosophically speaking, even a complete system of mathematical axioms is incapable of being proved by itself.

## Acknowledgement

This work is partially funded by CTI (Commission pour la Technologie et l¹ Innovation) and Natural Science Foundation of China (No. 60003012). The first author is financially supported by the ERCIM fellowship during his academic visit at EPFL.

## Reference

Allen J. F. 1983. Maintaining knowledge about temporal intervals. *Communications of the ACM*, 26(11), pp.832–843.

Bonissone Piero P. and Richard M. Tong(Eds.). 1985. Reasoning with uncertainty in expert systems. *International Journal of Man-Machine Studies*, 22(3), pp.241-250.

Bosc Patrick and Henri Prade. 1996. An introduction to the fuzzy set and possibility theory-based treatment of flexible queries and uncertain or imprecise databases. *Uncertainty Management in Information Systems*, Kluwer Academic Publishers, pp.285-324.

Clementini Eliseo, Paolino Di Felice, and Peter van Oosterom. 1993. A small set of formal topological relationships suitable for end-user interaction. *SSD*, pp.277-295.

Clementini Eliseo and Paolino Di Felice. 1996. An algebraic model for spatial objects with indeterminate boundaries. *Geographic Objects with Indeterminate Boundaries*, P. Burrough and A. Frank (Eds.), pp.155-169, Taylor & Francis.

Clifford James, Curtis E. Dyreson, Richard T. Snodgrass, Tomás Isakowitz, and Christian S. Jensen. 1995. ``Now''. *The TSQL2 Temporal Query Language*, pp.383-392.

Cohn A. G. and N. M. Gotts. 1996. The `egg-yolk' representation of regions with indeterminate boundaries. *Proceedings of GISDATA Specialist Meeting on Geographic Objects with Undetermined Boundaries*, P. Burrough and A M Frank (Eds.), pp.171-187, Francis Taylor.

Dyreson Curtis E. and Richard T. Snodgrass. 1993. Valid-time indeterminacy. *ICDE*, pp.335-343.

Guesgen Hans W. 2001. Fuzzifying spatial relations. *Applying Soft Computing in Defining Spatial Relations*, Pascal Matsakis and Les M. Sztandera (Eds.), Physica-Verlag, Heidelberg, Germany, pp.1-16.

Klir G. J. and Wierman. 1997. *Uncertainty-based information: elements of generalized information theory*. Creighton University, Omaha, Nebraska.

Parent C., Spaccapietra S., and Zimanyi E. 1999. Spatio-temporal conceptual models: data structures + space + time. *Proceedings of the 7th ACM Symposium on Advances in GIS*, Kansas City, Kansas.

Parent C., Spaccapietra S., and Zimanyi E. 2000. MurMur: database management of multiple representations. *Proceedings of AAAI-2000 Workshop on Spatial and Temporal Granularity,* Austin, Texas.

Schneider Markus. 1999. Uncertainty management for spatial data in databases: fuzzy spatial data types. *Proceedings of the 6th International Symposim on Advances in Spatial Databases (SSD),* LNCS 1651, Springer Verlag, pp.330-351.

Simon Parsons. 2001. *Qualitative methods for reasoning under uncertainty*, The MIT Press.

Smets Philippe. 1991. Varieties of ignorance and the need for well-founded theories. *Information Sciences*, 57, pp.135-144.

Smets Philippe. 1996. Imperfect information: imprecision and uncertainty. *Uncertainty Management in Information Systems*, Kluwer Academic Publishers, pp.225-254.

Smithson M.J. 1989. *Ignorance and uncertainty: emerging paradigms.* New York: Springer Verlag.

Zhang Jingxiong and Michael F. Goodchild. 2002. *Uncertainty in geographic information,* Taylor & Francis.

Http://www.fgdc.gov/metadata/contstan.html. 1998. *Content standard for digital geospatial metadata (CSDGM)*, Federal Geographic Data Committee. (CSDGM, 1998)

Http://lbdwww.epfl.ch/e/research/mads/#model. 1997. *MADS: modeling of application data with spatio-temporal features*, A Report Prepared by Christine Parent, Stefano Spaccapietra, Christelle Vangenot, Corinne Plazanet, Pier Donini, Yves Dennebouy. (MADS, 1997)